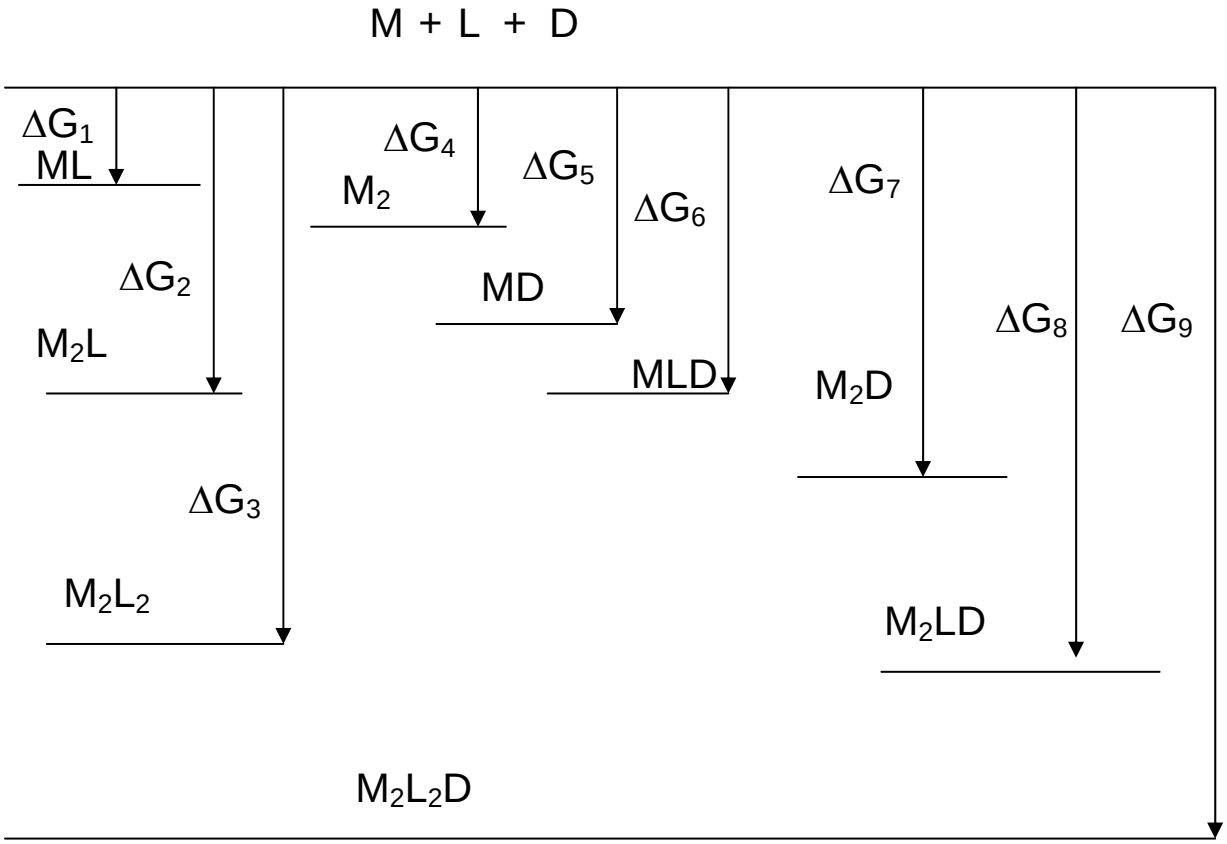
BIOEQS-Graphical User Interface USER MANUAL (ver. 1.2)

Catherine Royer, Joseph Beechem and Tilman Rosales

Introduction

Bioeqs is a data fitting software package designed for analysis of data for biomolecular interactions and conformational changes. The fitting algorithm is a Marquardt-Levenberg least squares algorithm, typical of many available data analysis routines. The uniqueness of BIOEQS lies in the solver, used to calculate the observable values from the input parameters. While most available programs use closed-form analytical expressions for the binding isotherms, BIOEQS uses a numerical algorithm based on constrained optimization with Lagrange multipliers to solve the set of **free energy** equations for a given model, in terms of the **species** concentration vector, given the mass balance constraints for the **elements**.



Any system can be described in terms of the elements that make it up. For example, for a protein that binds a small ligand and also DNA, the three fundamental elements are monomeric protein (M), ligand (L) and DNA (D). One can combine these three elements to form a variety of different complexes of varying stoichiometries. For example one can imagine monomer bound DNA (MD), DNA bound by liganded monomer (MLD), dimer bound DNA (M_2D), monomer bound by ligand (ML), unliganded dimer (M_2), and once (M_2L) and twice bound by ligand (M_2L_2), DNA bound by dimer with only one site liganded (M_2LD) and the full complex of fully liganded dimmer bound to DNA (M_2L_2D). This makes a total of

nine complexes, each having a corresponding free energy of formation from the three free elements. The total number of species is 12, corresponding to the 9 complexes plus the 3 free elements. Such a system is necessarily defined by only 9 free energy relations.

 # ΔG = # species - # elements

These 9 free energies can be defined in many ways. The BIOEQS convention is to define them always in terms of free energies of formation, for example the free energy of formation of the twice liganded dimer from free ligand and free monomer is ΔG_3 in the Figure below. The free energies for the transitions between species (for example that of binding of a second ligand to the once liganded dimer) can be calculated by subtracting ΔG_2 from ΔG_3.
The species are listed in matrix form

Species #	Species symbol	Stoichiometry Element 1	Stoichiometry Element 2	Stoichiometry Element 3
1	ML	1	1	0
2	M_2L	2	1	0
3	M_2L_2	2	2	0
4	M_2	2	0	0
5	MD	1	0	1
6	MLD	1	1	1
7	M_2D	2	0	1
8	M_2LD	2	1	1
9	M_2L_2D	2	2	1
10	M	1	0	0
11	L	0	1	0
12	D	0	0	1

The matrix is always terminated by an identity matrix with number of rows and columns equal to the number of elements. The maximum number of elements dimensioned for BIOEQS is three. The maximum number of species is 100.

There is a final column in the model file in the stoichiometric matrix that specifies the number of site isomers for each species. In the present example, one could imagine that the DNA half sites for each monomer may not be equivalent in sequence. Therefore the MD species and the MLD species may populate both of the site isomers corresponding to the two half sites. Their population will be weighted according to the free energies of formation of the two isomers. Unless one has an observable that actually is related to the differential population of each of the half sites (DNA footprinting results for example) it is best to consider that there is only one site isomer.

One also specifies on this page the concentration scaler for the elements 1, 2 and 3. Denaturant concentration is always in molar units.

Graphical User Interface Description

1.Model

The Graphical User Interface for BIOEQS (BIOEQS_GUI) consists of 4 pages. The first page is called **Model** and allows the user to input information about the model.

1. Number of elements
2. Number of species
3. Definition of element names (M for monomer, L for ligand, D for DNA for example)
4. The stoichiometry of each species
5. The number of site isomers for each species
6. The relative quantum yield of each species (the highest being 1)

If an answerfile already exists for the model to be used it can be loaded rather than recreated.

One must also specify the solver. This should be EQS for all systems except for monomer folding experiments by denaturant, temperature or pressure.

The second page called **Experiment** allows the user to specify how the experiment was done. The user must specify

1. What is the observable for the experiment
2. What kind of experiment is it
3. The concentration of the elements not being titrated
4. The denaturant concentration (if applicable, otherwise taken to be zero)
5. The pressure (if applicable, otherwise taken to be atmospheric)
6. The temperature in Kelvin
7. The data file name

2.Observable

There are 6 possible types of observables allowed in BIOEQS. The first three obspr, obsl1 and obsl2 correspond to observable values that are the result of the weighted average of the values for all of the species containing element 1 (usually taken to be protein) for obspr, element2 (usually taken to be the first ligand) for obsl1, and element 3 (usually taken to be the second ligand) for obsl2. The observable value at each data point in the experiment is the weighted average for the fractional population of each of the species at that data point (given the free energy parameters and concentrations) and for the relative quantum yield of each species (Q=1 for all species by default).

The second type of observable (xprot, xfrl and xfrl2) correspond to the fractional population of a particular species with respect to total element 1= protein (xprot), total element 2= ligand 1 (xfrl1) or total element 3= ligand 2. One must of course have a physical observable that reports specifically on this species. If one chooses xfrl1 for example, then after that one must specify which of the ligand 1 (element2) containing species is being designated (xfrl1 03) for example would correspond to twice liganded dimer (M_2L_2) in our

example. Then one must also specify which site isomer is being reported on by the observable. In most cases there will only be one, so one puts in xfrl1 03 01.

3.Experiment type.

BIOEQS supports 6 experiment types, the first three 1st element, 2nd element and 3rd element correspond to titrations by element 1 (protein), element 2 (ligand 1) and element 3. In these cases the first column in the data file must correspond to the different concentration points of the titration. The other three types of experiments are thermodynamic in nature. They are Folding, in which the first column of the data file corresponds to the different denaturant concentration points (in Molar units), Pressure, in which the firs column of the data file corresponds to the pressure points (in bar) and Temperature in which the firs column of the data file corresponds to the temperature points (in Kelvin).
After that one must specify the constant conditions of the experiment (i.e. the concentraiton of the elements that do not change during the experiment, and the thermodynamic conditions that do not change (denaturant concentration = 0 by default, pressure = atmospheric by default and temperature in Kelvin. Finally the data filename for each experiment must be entered.

Experimental Parameters

The third page **Experimental Parameters,** is where the fitting parameter initial guesses are entered. Depending upon the type of experiment there are different parameters. First there are the thermodynamic parameters. For titrations there will be one ΔG of formation per complex. For denaturant folding experiments there will be a ΔG and an m-value (of formation) per complex, for pressure experiments there will be a ΔG and a ΔV of formation per complex and for temperature experiments there will be a ΔG and a ΔH of formation per complex. (We note that the temperature experiments have not been debugged for the user interface version of BIOEQS, so use at your own risk!)

Note that in the example below, the model has been proposed such that protein binding to DNA is cooperative (ΔG7 > than 2 x ΔG5) but DNA binding disfavors ligation (ΔG6 – ΔG5 < ΔG1). This is carried through for all of the free energies of complex formation.

The second type of fitting parameter corresponds to those associated with the observables. If the observable is of the type weighted average of one of the elements (obspr, obsl1 or obsl2) then each species (obsl2 in our example) must be given a value (Xval) that corresponds to the value that would be observed if 100% of the element is in that species. For species that do not contain the element the value should be fixed at zero.

In our example, the observable is a the fluorescence anisotropy of a dye covalently attached to the DNA (element 3 = ligand 2), so the observable is obsl2. All species that do not contain DNA (M, ML, M_2, M_2L, M_2L_2) should have their Xvals set to 0 and fixed.

In our example we set the free DNA to a number close to what is observed at the beginning of the titration (75 milli-anisotropy units), and that for the dimer bound DNA to a value close to that observed at the end of the titration. Unless there are obvious bumps in the curve, it is best

to fix the intermediates somewhere in between. In the example below, it turned out that all of the anisotropy values of the DNA bound by liganded protein were a bit lower than in absence of ligand and so this is also apparent in the Xval values specified for these species.

If the observable corresponds to the fractional population of a particular species then only two parameters are given, the value observed when the species fractional population is 0% and that observed when it is 100%. These numbers are called Sval (starting value) and Eval (ending value).

In all cases the observable can be any type of numerical value (i.e., radioactive counts on a gel, fluorescence anisotropy, ellipticity, intensity of any spectroscopic signal) Only raw intensity values need not be corrected for quantum yield differences (if they exist).

Answerfile example

```
Total number of experiments = 2
 Data type:TIMEGS
 Filename:C:\data\cgfbp2.dat
 -------parameter type-  M  L  D---logic #---  ---initial value---
   model.ans deltaG              01  01      5.0000000E+00
   model.ans deltaG              02  02      1.2000000E+01
   model.ans deltaG              03  03      1.7000000E+01
   model.ans deltaG              04  04      7.0000000E+00
   model.ans deltaG              05  05      9.0000000E+00
   model.ans deltaG              06  06      1.3000000E+01
   model.ans deltaG              07  07      2.0000000E+01
   model.ans deltaG              08  08      2.4000000E+01
   model.ans deltaG              09  09      2.7000000E+01
   model.ans Xval   fix          10  10      0.0000000E+00
   model.ans Xval   fix          11  11      0.0000000E+00
   model.ans Xval   fix          12  12      0.0000000E+00
   model.ans Xval   fix          13  13      0.0000000E+00
   model.ans Xval   fix          14  14      8.5000000E+01
   model.ans Xval   fix          15  15      8.2000000E+01
   model.ans Xval               16  16      9.7000000E+01
   model.ans Xval   fix          17  17      9.3000000E+01
   model.ans Xval               18  18      9.0000000E+01
   model.ans Xval   fix          19  19      0.0000000E+00
   model.ans Xval   fix          20  20      0.0000000E+00
   model.ans Xval               21  21      7.5000000E+01
 ---------------------------------------------------------------
 Data type:TIMEGS
 Filename:C:\data\cgnfs2.dat
 -------parameter type-  M  L  D---logic #---  ---initial value---
   model.ans deltaG              22  22
   model.ans deltaG              23  23
   model.ans deltaG              24  24
   model.ans deltaG              25  25
   model.ans deltaG              26  26
```

```
 model.ans deltaG          27  27
 model.ans deltaG          28  28
 model.ans deltaG          29  29
 model.ans deltaG          30  30
 model.ans Xval            31  31
 model.ans Xval            32  32
 model.ans Xval            33  33
 model.ans Xval            34  34
 model.ans Xval            35  35
 model.ans Xval            36  36
 model.ans Xval            37  37
 model.ans Xval            38  38
 model.ans Xval            39  39
 model.ans Xval            40  40
 model.ans Xval            41  41
 model.ans Xval            42  42
---------------------------------------------------------------
grstate parameters follow:
obsl2 00 00 prot  0000.000000 2000.000000 0000.001000 0000.000000 0000.000000
291.00
obsl1 00 00 prot  0000.000000 0050.000000 0000.001000 0000.000000 0000.000000
291.00
```

Above is a copy of the ascii answerfile generated by the graphic user interface of BIOEQS for the model specified above; with two titrations at two different concentrations of the second element. Below is a copy of the answerfile corresponding to the stoichiometric model that is embedded in the global answerfile and called model.ans.

model.ans file example

```
Number of Free Energy Equations:9
Number of Possible Species:12
Number of Elements:3
Concentration Scaler(3=mM,6=uM,9=nM):6
Solver:EQS
*********************************************************************************
          01(01.00)(01.00)(00.00)(01.00)
          02(02.00)(01.00)(00.00)(01.00)
          03(02.00)(02.00)(00.00)(01.00)
          04(02.00)(00.00)(00.00)(01.00)      This is the stoichiometric matrix with the
          05(01.00)(00.00)(01.00)(01.00)       number of site isomers in the last column
          06(01.00)(01.00)(01.00)(01.00)
          07(02.00)(00.00)(01.00)(01.00)
          08(02.00)(01.00)(01.00)(01.00)
          09(02.00)(02.00)(01.00)(01.00)
          10(01.00)(00.00)(00.00)(01.00)     Note here the identity matrix for the elements
          11(00.00)(01.00)(00.00)(01.00)      in the first 3 colmuns and last 3 lines
          12(00.00)(00.00)(01.00)(01.00)
*********************************************************************************
```

01(0.00)(0.00)(0.00)
02(0.00)(0.00)(0.00) This matrix is no longer used
03(0.00)(0.00)(0.00)
04(0.00)(0.00)(0.00)
05(0.00)(0.00)(0.00)
06(0.00)(0.00)(0.00)
07(0.00)(0.00)(0.00)
08(0.00)(0.00)(0.00)
09(0.00)(0.00)(0.00)
10(0.00)(0.00)(0.00)
11(0.00)(0.00)(0.00)
12(0.00)(0.00)(0.00)

****

01(1.00) This last vector is the quantum yield vector
02(1.00)
03(1.00)
04(1.00)
05(1.00)
06(1.00)
07(1.00)
08(1.00)
09(1.00)
10(1.00)
11(1.00)
12(1.00)

Below are examples of the two data files to be analyzed. The first column corresponds to the different concentrations of the element being titrated into the solution (in this case element 1 = protein). The second column corresponds to the value observed experimentally at that concentration and the third column is the estimated uncertainty on that value. Note that the formatting of the numbers of the two data files is not exactly the same. The only things tat are important are

1. that the first column be expressed in the units specified on the first page (micromolar in our case such that the titration was done between 0.1 pM and 1 micromolar)
2. All of the numbers must be real (i.e. have a decimal point).

cgfbp2.dat

0.00010	75.0	2.0
0.00030	76.5	2.0
0.00050	76.5	2.0
0.00075	77.0	2.0
0.001	77.5	2.0
0.0015	77.5	2.0
0.002	78.0	2.0
0.003	79.0	2.0
0.004	79.5	2.0
0.006	80.5	2.0
0.007	82.0	2.0

0.0085	83.5	2.0
0.012	84.5	2.0
0.018	86.0	2.0
0.027	88.0	2.0
0.036	88.5	2.0
0.054	90.0	2.0
0.072	93.5	2.0
0.1	94.0	2.0
0.14	94.5	2.0
0.18	93.5	2.0
0.3	95.0	2.0
0.5	95.5	2.0
0.8	97.5	2.0
1.00	96.5	2.0

cgnfs2.dat

0.100000E-03	75.7046	2.000000
0.300000E-03	75.9593	2.000000
0.500000E-03	76.2580	2.000000
0.750000E-03	76.6791	2.000000
0.100000E-02	77.1407	2.000000
0.150000E-02	78.1472	2.000000
0.200000E-02	79.2214	2.000000
0.300000E-02	81.4406	2.000000
0.420000E-02	84.0421	2.000000
0.600000E-02	87.5135	2.000000
0.700000E-02	89.1448	2.000000
0.850000E-02	91.1752	2.000000
0.121000E-01	94.3463	2.000000
0.180000E-01	96.6444	2.000000
0.270000E-01	97.8643	2.000000
0.360000E-01	98.3334	2.000000
0.540000E-01	98.7077	2.000000
0.720000E-01	98.8609	2.000000
0.100000	98.9730	2.000000
0.140000	99.0462	2.000000
0.180000	99.0837	2.000000
0.300000	99.1323	2.000000
0.500000	99.1594	2.000000
0.800000	99.1740	2.000000
1.00000	99.1788	2.000000

One should save the answerfiles on the previous pages into any folder on the computer. BIOEQS_GUI makes another temporary copy that it uses to run the program.

4. Results

By clicking the "View and copy Results…" button in the Experimental Parameters page, the program will copy the results of the bioeqsg.dll to the same location as the answerfiles and will take you to the **Results** page. All of the results (parameters file as well as a file with data, fit and residuals) are saved into ascii output files that are stored in the same location as you put your answerfiles. These files have the same name as the answerfile, but instead of an extension called .ans they have extensions of .res and .tmp, respectively. The returned values for all experiments are shown on this page including the global chi-square, chi-squares for all the experiments and pertinent information. The experimental fits are graphed by pressing the "Graph Experimental Fits…" button.

Covariance

Click on the "covariance" button in the Results tab to access a page where you are able to perform rigorous confidence limit testing of the recovered parameters. You can do up to 20 different parameters at a time. The number of "points" per fitting parameter has been set to 100. The analysis requires as input the fitting parameter (selected from a drop down list), the starting and ending chi-square values and the interval desired. After pressing the "Analyze" button, the returned values are displayed in the output table. (The amount required for the calculation will vary depending on the total number of variables). A graph of the returned chi-square vs fitting parameter is also displayed. A blue and a green line may also be plotted showing the 0.67 and the 0.9772 confidence level for the chosen experiment. If the 0.67 or the 0.9772 confidence levels are outside the graphing range, an "OUT OF BOUNDS" will be displayed color coded to the appropriate confidence level.

5. Simulations

The simulation options can be accessed by checking the simulation box on the model page. Simulations are performed in a similar manner as the normal analysis. You can add noise to the simulations by giving a seed for the noise generator and the percent noise. No noise is added if the seed is set to zero.

The files required for the simulation can be created in the experiment page or the users can simply use their own files. The user can input the filename for the experiment, the starting and ending value (when needed) and the number of points desired. For titrations, the step size is 0.1 log units. The step size for the other experiment types is a linear increment given by (End Value –Start Value)/ No. of points. Pressing "create files" will create the experiment file at the selected location with a ".dat" extension.

The program requires the same input as the regular analysis. When all the variables are filled in (i.e. "Master Table"), press "Calculate" to create the simulations files. Pressing the "View and copy results…" button takes you to the Results page where you can also graph the simulated experiments by pressing the "Graph experimental fits" button.

General warnings

The regional settings on the computer need to have the decimal place indicator set as a period and not a comma. THIS IS VERY IMPORTANT. Every time the program runs, it will attempt to change the regional settings. When the program is closed by pressing the EXIT button, it will return the settings to their original values. If the program crashes or is terminated, these settings will not be restored. You can always change them back by going to the Control Panel → Regional and Language Settings and selecting "Customize". Here the decimal descriptor and the digit grouping descriptor can be changed.

It is recommended that any previous versions of this program be removed from the user's computer before installing a newer version of the program.

Problems?

Surely there will be some. Do not hesitate to e-mail catherine.royer@cbs.cnrs.fr

We will try to fix them.

Model Making Tool "HOW TO"

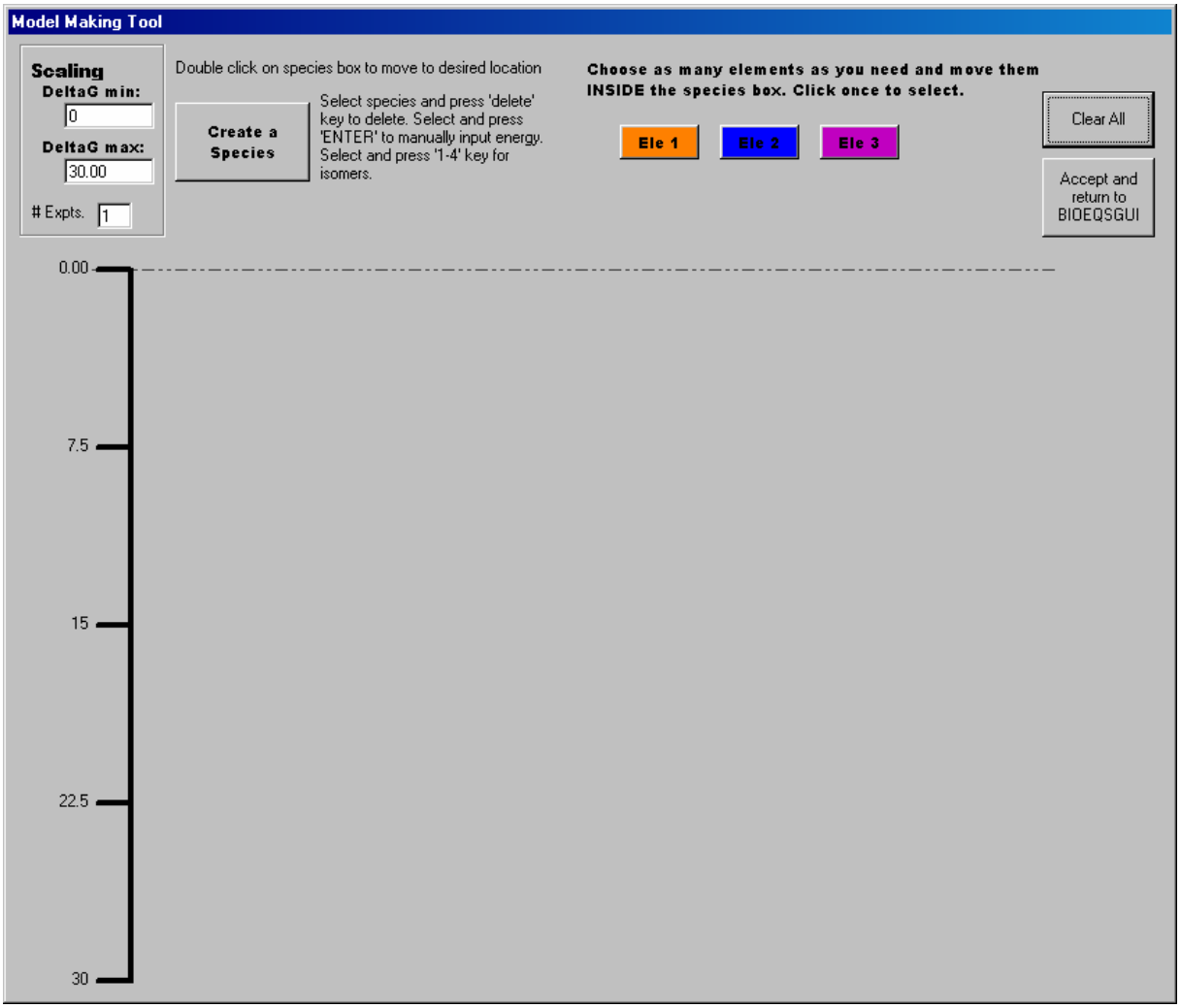
Version 1.00 Date: 5/21/07

When you first load the "model making tool window", you should have a clean slate.
On the left hand side you can choose the range of DeltaG, you want to use. You need to simply type in the minimum and the maximum value on the appropriate boxes, because the area is being *scaled* to these values.
Below the DeltaG max box is the # of experiments you plan to analyze or simulate. Type in the number of experiments you plan to analyze or simulate. You will be able to change this number on the main form if you need to. This option will fill out the main form # of experiments also.

Basic interface principles:

As you get used to the interface, you will be able to move around and do things differently. I recommend creating the species that you need *first*, then filling them with the appropriate elements corresponding to your experiment.
The basic interface uses "single click" mostly. One exception is a double click done on the species box to drag the species box around. Note that the value of the energy displayed on the species box as it is dragged corresponds to that of the top part of the box.

Fig. 1.The startup form for the model making tool.



To create a “Species”

Click on the button “Create a Species”. A “box” will appear right below this button.

You have *three* options to move the species box around:

1. Double click the box to drag it around any where you want. Be careful not to drag it out of the form space. (It can happen…). Once the species box is no longer visible, you will have to push the “Clear All” and start anew.

2. Doing a single click will "select" the species box. Note that at the time there is no apparent change to the look of the species box you selected.

You can move a species box up or down using the PAGE UP and PAGE DOWN buttons on your keyboard after "selecting the species".

You can also press "ALT" + any arrow key (LEFT, UP, DOWN and RIGHT) to move the species around. (Note that if you have any "elements" inside the box, they will move with the box *only* if you use the single click option).

3.The last option is to single click the species box and then press "ENTER", this will bring up a text box where you can type the exact energy you want. After typing the exact DeltaG you want, press ENTER. The box will move to the exact energy you want. Notice that the horizontal position is preserved!

To delete a species simply select the species and hit "delete" key.

To create an element:

Click on the Element box to create an element, you'll notice that there are three different elements each with a different color and shape. You should always start using the orange circle since it corresponds to element 1 on the tabbed forms. In the same manner, the blue square corresponds to element 2 and the triangle to element 3.

Creating the first element also creates an image on the form to remind you that you have an element of that type somewhere on the form. Again be careful not to accidentally drag an icon outside the boundaries (A check to keep them inbounds is on the works)

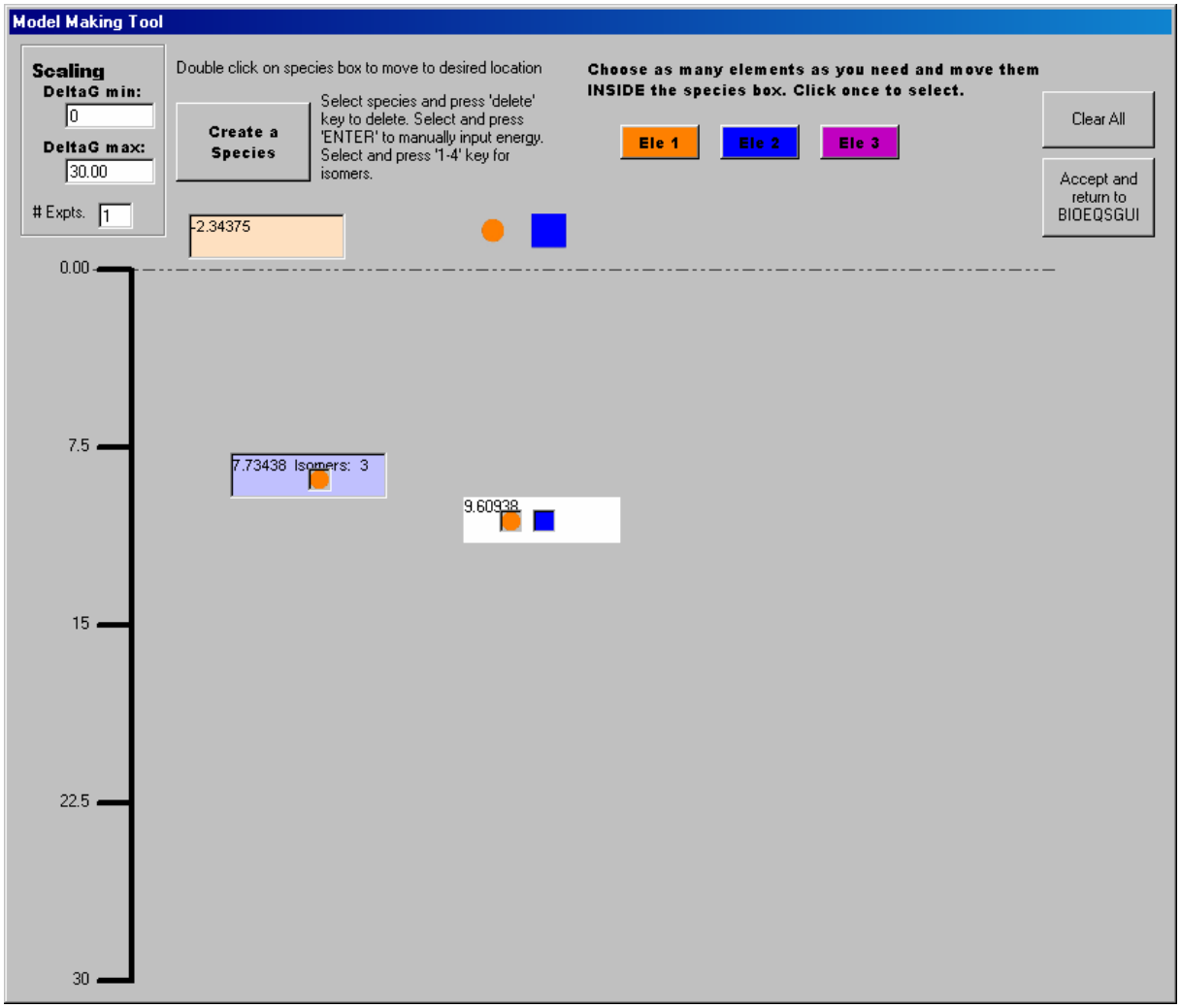
To pick up an element, do a single click on the element icon. This attaches the icon to the mouse movement. Do a single click anywhere on the form to drop the element icon. For the program to count the element, it has to be inside a species box! The program simply checks what's inside the boundaries of the species box (top, bottom and sides). **This is very important, the element icons need to be inside the species box!**

To delete an element simply select the species and press "delete" key

To create an isomer:

At this time, the number of isomers is done by selecting the species box (with one single click) and pressing the number key on the keyboard 1 through 4 (Only up to four isomers allowed per element. This is a GUI limitation) The number of isomers for that species is displayed on the box itself.

Fig 2. NOTE: A species box changes color when double clicked ; but not when single clicked. The box becomes white after hitting ENTER so that you can type in the exact energy you want. Remember to press ENTER when done.



If you get into trouble, you can always delete all by pressing the “Clear All” button.